313 Assignment 4

Ben Walsh

2023-05-25

```
library(sm)
```

```
## Package 'sm', version 2.2-5.7: type help(sm) for summary information
```

```
library(ks)
```

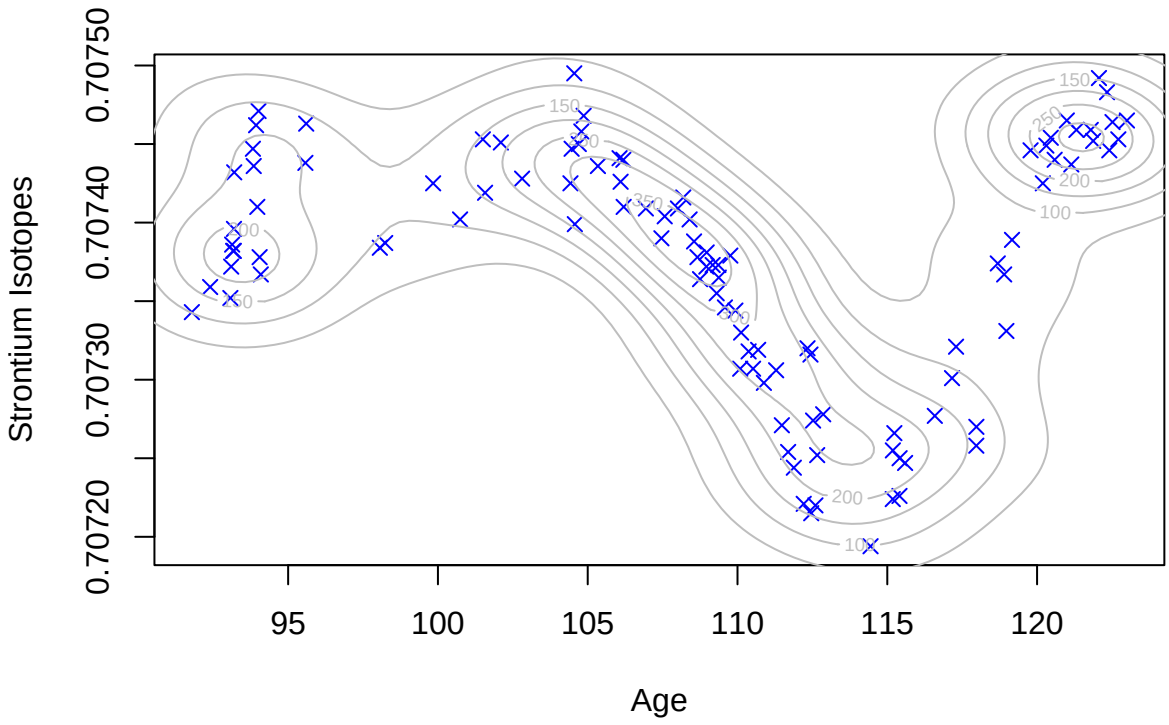
```
##
## Attaching package: 'ks'
```

```
## The following object is masked from 'package:sm':
##
##     binning
```

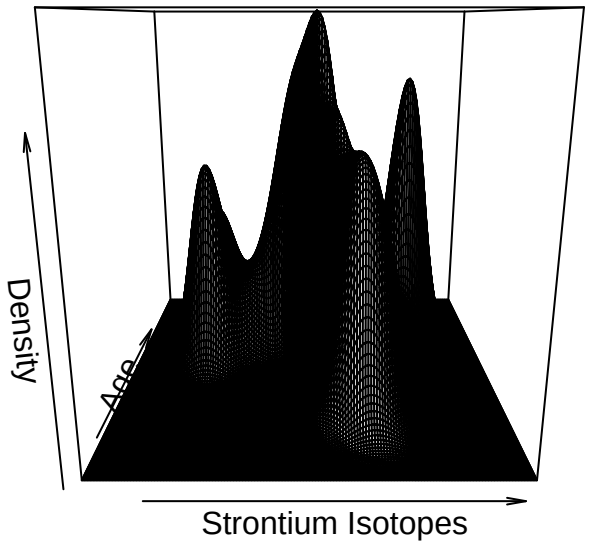
Problem 1 Multivariate Estimation

```
#Question 1a)
data = read.table("fossil.txt", header = TRUE)
D = kde(data)
#hashed so that they dont run in Rmarkdown:
#D$eval.points[[1]]
#D$eval.points[[2]]
#D$estimate

#Question 1b)
plot(data, pch=4, xlab="Age", ylab="Strontium Isotopes", col="blue")
contour(D$eval.points[[1]], D$eval.points[[2]], z=D$estimate, col="gray",
        add = TRUE) #no bandwidth selected = data driven selection
```



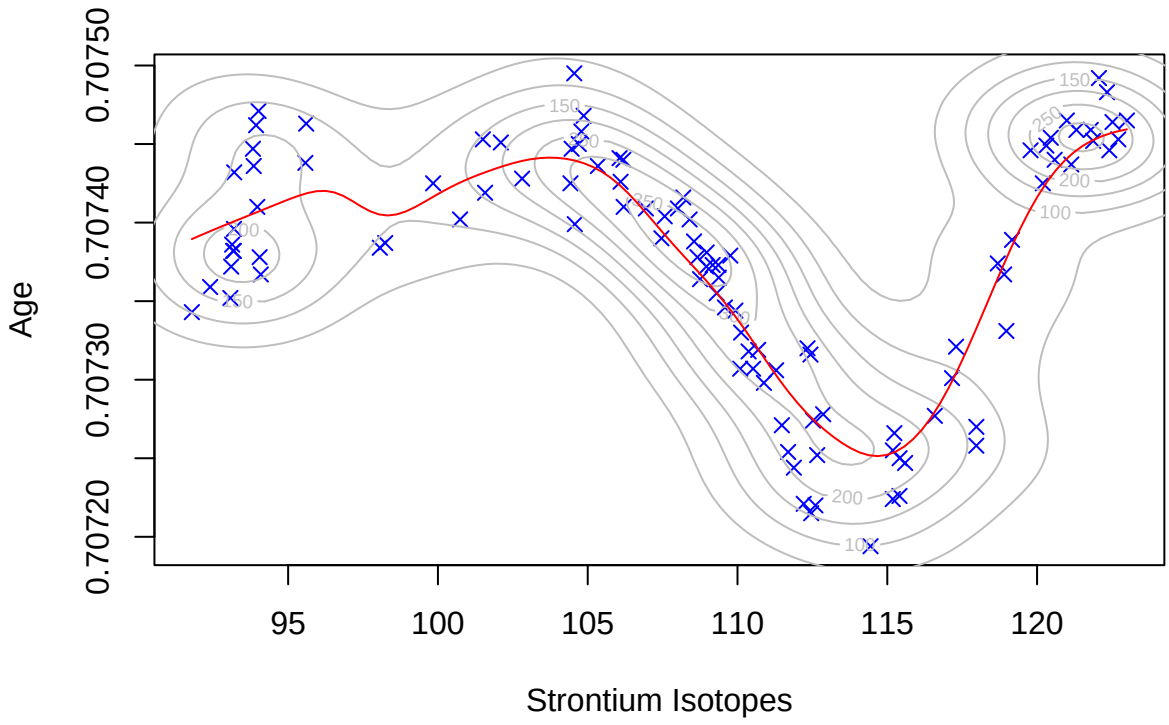

```
persp(D$eval.points[[1]], D$eval.points[[2]], D$estimate, xlab='Strontium Isotopes',
      ylab='Age', zlab='Density')
```




```
#Question 1c)
model = ksmooth(data[,1],data[,2], 'normal', bandwidth=4)
```

I chose the bandwidth of 4 as I believe it produces the most ideal regression line, without over fitting or over smoothing. It still displays the left and right peaks and still includes strength of the middle trough.

```
#Question 1d)
plot(data, pch=4, xlab="Strontium Isotopes", ylab="Age", col="blue")
contour(D$eval.points[[1]], D$eval.points[[2]], z=D$estimate, col="gray",
        add = TRUE) #no bandwidth selected = data driven selection
lines(ksmooth(data[,1],data[,2], 'normal', bandwidth=4), col='red')
```

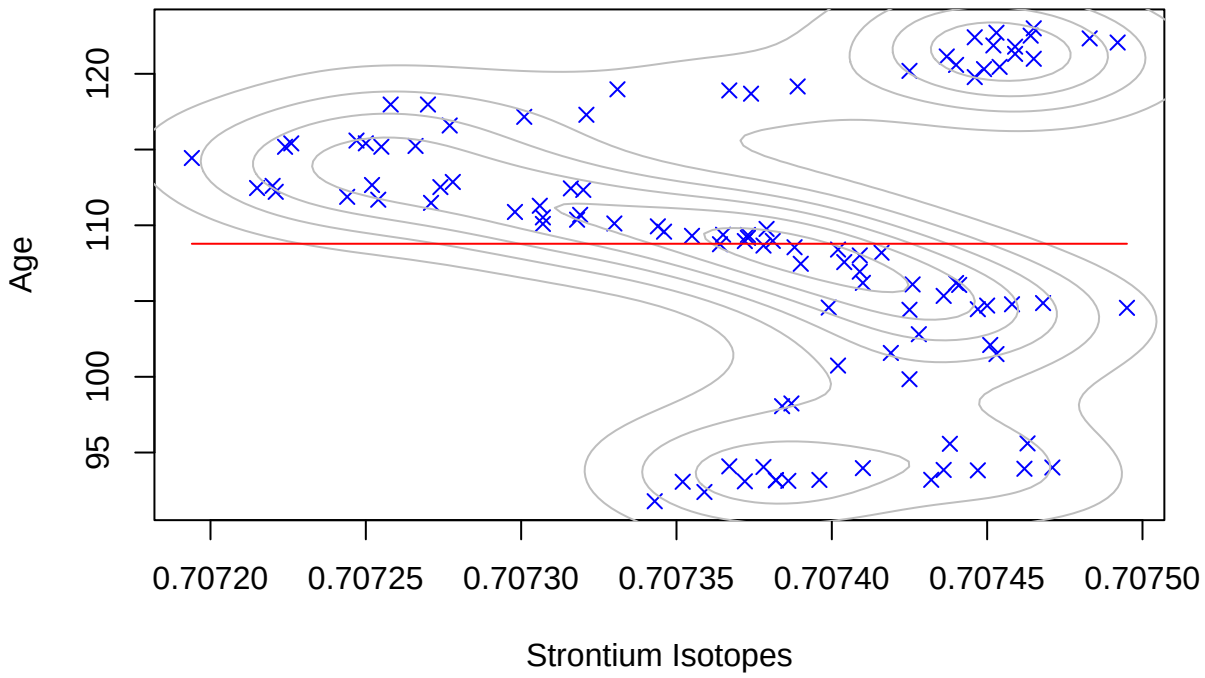



```
#Question 1e)
data_new = data[c("strontium.ratio", "age")]
D = kde(data_new)
#D$estimate

#Question 1f)
model2 = ksmooth(data_new[,1],data_new[,2], 'normal', bandwidth=4)
```

Here the bandwidth choice is somewhat irrelevant, the Nadaraya-Watson estimator is not ideal here, however I have chosen a bandwidth of 1. No matter the bandwidth, there never seems to be a good regression function.

```
#Question 1g)
plot(data_new, pch=4, xlab="Strontium Isotopes", ylab="Age", col="blue")
contour(D$eval.points[[1]], D$eval.points[[2]], z=D$estimate, col="gray",
        add = TRUE) #no bandwidth selected = data driven selection
lines(ksmooth(data_new[,1],data_new[,2], 'normal', bandwidth=1), col='red')
```



Reversing the x and y axis produces the model above. Here a straight line is produced for our Nadaraya-Watson estimator, this indicates that our estimator is oversmoothed. We could make a mi-nute bandwidth for the scale of the Strontium Isotopes, however it is obvious that this model is just not good for this dataset (data_new).

Problem 2 Interpretation

2a) Both density estimates are similar in shape, both with 4 regions of protrusion depicting a quadmodel shape (in 3D). Each peak together produces an s shape. Reversing the axis flipped the second density estimates, hence the very similar shapes. The only reason they look different is because they have been stretched x ways. We can get one density from the other without looking at the data again as both contour plots have the same shape, If we were to flip the axis back we would get the same density.

2b)

The regression function for the density with Strontium Isotopes as the response produces an accurate estimation. The trend follows the contour nicely, and with a bandwidth of 4 the function is not over smoothed nor over fitted. The other regression function however struggles to produce an accurate trend line. The function is a linear line and no matter the bandwidth we cannot produce a good estimator. You cannot get one regression function from the other as the estimator/trend line is completely different, unlike the contour, where the shape and distribution across both x and y is the same.

2c)

It would be a great for predicting the ratios of Strontium Isotopes from the age of fossil as our regression function follows our contour plot nicely, producing an accurate and fair estimator.

2d)

Determining the Age of fossil from the ratios of Strontium Isotopes would work horribly. the regression function produces a horizontal linear line, coming no where close to the shape of the contour.